

Chetan Aitaraju

Email: chetanaitaraju2023@gmail.com

Contact: 3475428533

<https://www.linkedin.com/in/chetan-aitaraju-093a6b1b2/>

PROFESSIONAL SUMMARY:

- Around 4+ years of extensive experience in Information Technology with expertise on Data Analytics, Data Design, Development, Implementation, Testing and Deployment of Software Applications in Finance, Insurance, domains.
- Working experience on designing and implementation complete end to end Hadoop infrastructure using HDFS, MapReduce, Hive, HBase, Kafka, Sqoop, Spark, No SQL, Postman, and Python
- Created Data Frames and performed analysis using Spark SQL.
- Acute knowledge on Spark Streaming and Spark Machine Learning Libraries.
- Experienced in writing the automatic scripts for monitoring the file systems, key MapR services.
- Implemented continuous integration & deployment (CICD) through Jenkins for Hadoop jobs.
- Good Knowledge on Cloudera distributions and in Amazon simple storage service (Amazon S3), AWS Redshift, Lambda and Amazon EC2, Amazon EMR.
- Performed transformations on the imported data and exported back to RDBMS.
- Worked on Amazon Web service (AWS) to integrate EMR with Spark 2 and S3 storage and Snowflake.
- Experience in writing queries in HQL (Hive Query Language), to perform data analysis.
- Created Hive External and Managed Tables.
- Implemented Partitioning and Bucketing on Hive tables for Hive Query Optimization.
- Integrated Flume with Kafka, using Flume both as a producer and consumer (concept of FLAFKA).
- Good Exposure to create various dashboard in Reporting Tools like SAS, Tableau, Power BI, BO, QlikView used various filters, sets while dealing with huge volume of data.
- Experience in various Database such as Oracle, Teradata, Informix and DB2.
- Experience with NoSQL like MongoDB, HBase and PostgreSQL like Greenplum

PROFESSIONAL EXPERIENCE

Skillo Inc, Charlotte NC

March 2023 to Present

Data Engineer

Responsibilities:

- Implemented Installation and configuration of multi-node cluster on Cloud using Amazon Web Services (AWS) on EC2.
- Handled AWS Management Tools as Cloud watch and Cloud Trail.
- Stored the log files in AWS S3. Used versioning in S3 buckets where the highly sensitive information is stored.
- Integrated AWS Dynamo DB using AWS lambda to store the values of items and backup the DynamoDB streams.
- Automated Regular AWS tasks like snapshots creation using Python scripts.
- Designed data warehouses on platforms such as AWS Redshift, Azure SQL Data Warehouse, and other high-performance platforms.
- Install and configure Apache Airflow for AWS S3 bucket and created dags to run the Airflow.
- Prepared scripts to automate the ingestion process using Pyspark and Scala as needed through various sources such as API, AWS S3, Teradata and Redshift.
- Created multiple scripts to automate ETL/ ELT process using Pyspark from multiple sources.
- Developed Pyspark scripts utilizing SQL and RDD in spark for data analysis and storing back into S3.
- Developed code in Spark SQL for implementing Business logic with python as programming language.
- Worked on Sequence files, Map side joins, bucketing, partitioning for hive performance enhancement and storage improvement.
- Used Spark Streaming to divide streaming data into batches as an input to Spark engine for batch processing.
- Managed multiple Kubernetes clusters in a production environment.
- Wrote Spark applications for data validation, cleansing, transformation, and custom aggregation and used Spark engine, Spark SQL for data analysis and provided to the data scientists for further analysis.

- Data Integrity checks has been handled using hive queries, Hadoop, and Spark.

Environment: AWS, JMeter, Kafka, Ansible, Jenkins, Docker, Linux, GIT, Cloud Watch, Python, Shell Scripting, Golang, Web Sphere, Splunk, Soap UI, Kubernetes, Terraform, PowerShell.

Cuspyd, Hyderabad

March 2019 to Jan 2022

Tech Mahindra

Data Engineer

Responsibilities

- Involved in complete Implementation lifecycle, specialized in writing custom MapReduce, and Hive
- Extensively used Hive/HQL or Hive queries to query or search for a string in Hive tables in HDFS
- Continuous monitoring and managing the Hadoop cluster using Cloudera Manager
- Implemented Spark using Python and Spark SQL for faster processing of data.
- Dumped the data from HDFS to Oracle database and vice-versa using Sqoop.
- Extensively involved in Installation and configuration of Cloudera Hadoop Distribution.
- Provided support for EBS, Trusted Advisor, Cloud Watch, Cloud Front, IAM, Security Groups, Auto-Scaling, AWS CLI and Cloud Watch Monitoring creation and update.
- Worked with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism.
- Deployed Lambda and other dependencies into AWS to automate EMR Spin for Data Lake jobs.
- Scheduled spark applications/Steps in AWS EMR cluster.
- Extensively used event-driven and scheduled AWS Lambda functions to trigger various AWS resources.
- Developed spark applications for performing large scale transformations and denormalization of relational datasets.
- Developed and executed a migration strategy to move Data Warehouse from SAP to AWS Redshift.
- Performed Impact Analysis of the changes done to the existing mappings and provided the feedback.
- Create mappings using reusable components like worklets, mapplets using other reusable transformations.
- Participated in providing the project estimates for development team efforts for the offshore as well as on-site.
- Coordinated and monitored the project progress to ensure the timely flow and complete delivery of the project.
- Worked on Informatica Source Analyzer, Mapping Designer & Mapplet, and Transformations.

Environment: Hadoop, HDFS, Hive, MapReduce, Impala, Sqoop, SQL, Informatica, Python, Flume, Spark, Yarn, Pig, Oozie, Linux, AWS, Tableau, Maven, Jenkins, Autosys, Oracle, Sql Server, Sql, Teradata.

TECHNICAL SKILLS:

Big Data Ecosystem: HDFS, MapReduce, HBase, Pig, Hive, Sqoop, Kafka Flume, Cassandra, Impala, Oozie, Zookeeper, MapR, Amazon Web Services (AWS), EMR

Cloud Technologies: AWS, Azure, Google cloud platform (GCP)

Databases: Oracle 11g/10g/9i, MySQL, DB2, MS SQL Server, HBASE

Programming: Query Languages Java, SQL, Python Programming (Pandas, NumPy, SciPy), NoSQL, PySpark, Scala.

Data Engineer: Big Data Tools / Cloud / Visualization / Other Tools Databricks, Hadoop Distributed File System (HDFS), Hive, Pig, Sqoop, MapReduce, Spring Boot, Flume, Azure Databricks, Azure Data Explorer, Linux, PuTTY, Bash Shell, Unix, etc., Tableau, Power BI, SAS, We Intelligence, Crystal Reports, Dashboard Design.

EDUCATION DETAILS:

Master of science- MS | Cumberland University

Jan 2022 – April 2024

Bachelor's in computer | Pacific institute of Engineering Technology | New Delhi

May 2015-May 2019

CERTIFICATIONS:

AWS certified Devops Engineer- Professional | Amazon Web Services (AWS)

AWS certified Developer- Associate | Amazon Web Services (AWS)